Copyright

by

Jessica Sueji Kim

2020

The Report Committee for Jessica Sueji Kim
Certifies that this is the approved version of the following Report:

Improving Universal Screening Practices with Posttest Probability

APPROVED BY
SUPERVISING COMMITTEE:

David Klingbeil, Supervisor

Jessica O'Bleness

Improving Universal Screening Practices with Posttest Probability

by

Jessica Sueji Kim

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

August 2020

Abstract

Improving Universal Screening Practices with Posttest Probability

Jessica Sueji Kim, MA

The University of Texas at Austin, 2020

Supervisor: David Klingbeil

There is a growing number of calls for schools to adopt principles of evidence-based assessment (EBA) to address the challenges in effectively using universal screening data to inform intervention decisions. Researchers have begun evaluating the use of a metric that is aligned with EBA principles, called posttest probability, in school-based settings. The purpose of this literature review is to examine research on the proposed use of posttest probabilities in schools and discuss different factors that are associated with academic achievement that may be useful in calculating actionable posttest probabilities. An example to illustrate how this information can be applied in schools as well as areas for future research are provided.

Table of Contents

List of Tables

List of Figures

Introduction

Universal screening is a critical aspect in the provision of targeted prevention and tiered intervention services (Glover & Albers, 2007). Schools conduct universal screening to identify students who may be at risk for failure and make data-based decisions on how the school can best address students' needs. Based on screening scores, schools determine which students need to be provided with additional support. Particularly in the elementary grades, universal screening is widely used in schools across the United States in reading and math.

As more schools implement Response-to-Intervention (RTI) and Multi-tiered Systems of Support (MTSS) models, attention on the use of screening in educational contexts has increased (Clemens et al., 2016). The recent research on universal screening has indicated the need for improvement of screening efficiency and accuracy—as such there have been a growing number of calls for schools to adopt principles of evidence-based assessment (EBA; Pendergast et al., 2018; VanDerHeyden & Burns, 2018) to address the challenges in effectively using screening data to inform intervention decisions.

Research across disciplines has demonstrated that actuarial decision-making methods outperform aided clinical judgment in accuracy and consistency (Pendergast et al., 2018). Accordingly, researchers have begun evaluating the use of EBA principles such as adopting posttest probability in school-based settings to evaluate screening measures and decisions. One of the key components of the EBA approach is identifying the base rate or prevalence of a diagnosis (Youngstrom et al., 2015). Schools do not typically deal with diagnoses but are often concerned with other characteristics such as risk status for academic problems. Base rates provide a starting point for determining the probability a student is at-risk absent of additional information (Youngstrom et al., 2015) while also taking into consideration the context of the learning environment. Combining base rates, which function as the pretest probability, with information on relevant risk factors and screener information can provide a more comprehensive, actionable metric called posttest probability. Calculation of a posttest probability requires the consideration of the risk factors related to the outcome of interest in conjunction with the pretest probability; therefore, the purpose of the literature review is to discuss predictors of student achievement that may be useful factors to utilize when determining posttest probability.

Current Screening Practices in Schools

Various studies have looked at the current screening practices in schools (e.g. Balu et al., 2015; Jenkins et al., 2013; Mellard et al., 2009). These studies have shown that the most commonly used screeners are curriculum-based measurements (CBM), published reading assessments, and district or state assessments, and they often use the published norm- or criterion-based cut scores to determine tier placement within the RTI model (Balu et al., 2015; Jenkins et al., 2013; Mellard et al., 2009). The next most frequently reported method for determining at-risk students is the use of percentages of the local student population (e.g. the lowest performing 25% of students are targeted for intervention; Mellard et al., 2009). Some schools do not use cut points at all but rather a convergence of data from assessments and teacher reports which relies on the clinical judgement of school personnel (Mellard et al., 2009). Each of these methods present issues regarding the accuracy and effectiveness of the decisions made.

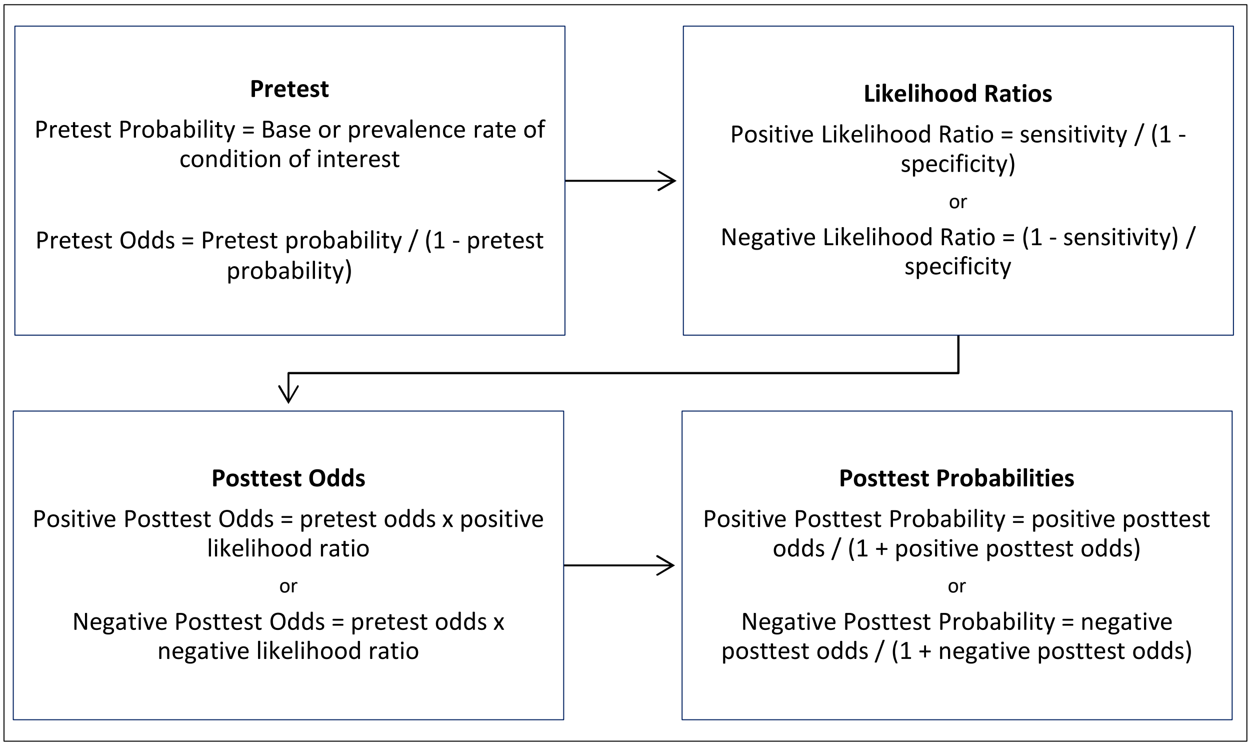
When screeners are used to identify students who are at-risk for academic failure based on norm- or criterion-referenced cut scores, the failure prevalence or base rates of the population of interest must match those of the norming sample, otherwise the predictive value of the cut score will not be the same. In addition, the diagnostic accuracy of a single screener, defined by the sensitivity and specificity, which are the percentage of true positive determinations and percentage of true negative determinations respectively, is usually not sufficient (Johnson et al., 2010). Similar issues arise with the method of targeting the lowest performing x percent of students. For schools that have very low or high base rates, this method could over- or under-identify students targeted for intervention. Thus, there is a push to implement a process that combines base rates (pretest probability), screener data and student characteristics that moderate achievement in the form of a diagnostic likelihood ratio to calculate a more accurate, actionable metric called a posttest probability.

Posttest Probability

Posttest probabilities can be positive or negative. A positive posttest probability (+PP) is the probability that a student who failed the screener will also fail the criterion, and a negative posttest probability (-PP) is the probability a student who has passed the screener will fail the criterion. These metrics are actionable in that a high +PP tells the decision maker that intervention should be provided to the student to prevent failure whereas a low -PP tells the decision maker that the student can be ruled out for intervention (VanDerHeyden, 2013). Calculating a posttest probability requires a combination of pretest probability and likelihood ratios by applying Bayes' theorem (see Figure 1).

Figure 1

Posttest Probability Calculations



Pretest probability is the probability that a condition is present before any additional information about risk factors or test data is known. This is usually equivalent to the local base or prevalence rate of the condition of interest. Pretest probability is used to compute the pretest odds (pretest probability/(1-pretest probability) which is then multiplied by a likelihood ratio to determine posttest odds. Likelihood ratios are ratios of probabilities that incorporate both sensitivity and specificity and like posttest probabilities, likelihood ratios can be either positive or negative. A positive likelihood ratio (sensitivity/(1-specificity)) tells the user how many times more likely a child who will fail the criterion is to have failed the screening whereas a negative likelihood ratio ((1-sensitivity)/specificity) indicates how many times less likely a child who will pass the criterion will have failed the screening (VanDerHeyden, 2013). A positive likelihood ratio is used to compute positive posttests odds and a negative ratio for negative posttest odds (posttest odds = pretest odds * likelihood ratio). Finally, the posttest odds would be used to calculate the posttest probability (posttest odds/(1 + posttest odds)).

Once posttest probabilities are calculated, they can be used within a threshold model to inform decision-making (Pauker & Kassirer, 1980). The posttest probability can fall into one of three zones: wait, test, or treat—if the posttest probability is higher than the test-treat threshold, the condition of interest is ruled in and becomes the focus in treatment planning. If the probability falls below the wait-test threshold, it is low enough where the condition can be considered ruled out, and if the probability falls between the two thresholds, more data should be gathered to revise the probability until it crosses either threshold (Youngstrom et al., 2015). VanDerHeyden (2013) suggested that posttest probabilities between 10 and 50 percent should be additionally assessed whereas posttest probabilities below 10 percent do not require intervention or further action, and posttest probabilities above 50 percent should receive intervention. VanDerHeyden (2013) also provided an example of how relevant risk factors are taken into consideration to determine differential treatment for patients with similar symptoms: consider a patient who is 25 years old, complains of a persistent cough, does not smoke, and has a sore throat and runny nose—characteristics like age and lifestyle suggest a low probability for a condition like lung cancer and would fall below the wait-test threshold; whereas for a 65-year-old patient also experiencing a persistent cough with a history of heavy smoking, the patient would be treated differently due to elevated base rates for a more serious diagnosis in populations of heavy smokers, and increased risk due to age. This example illustrates the utility of incorporating multiple factors to determine a posttest probability which in turn can be used to inform treatment decisions.

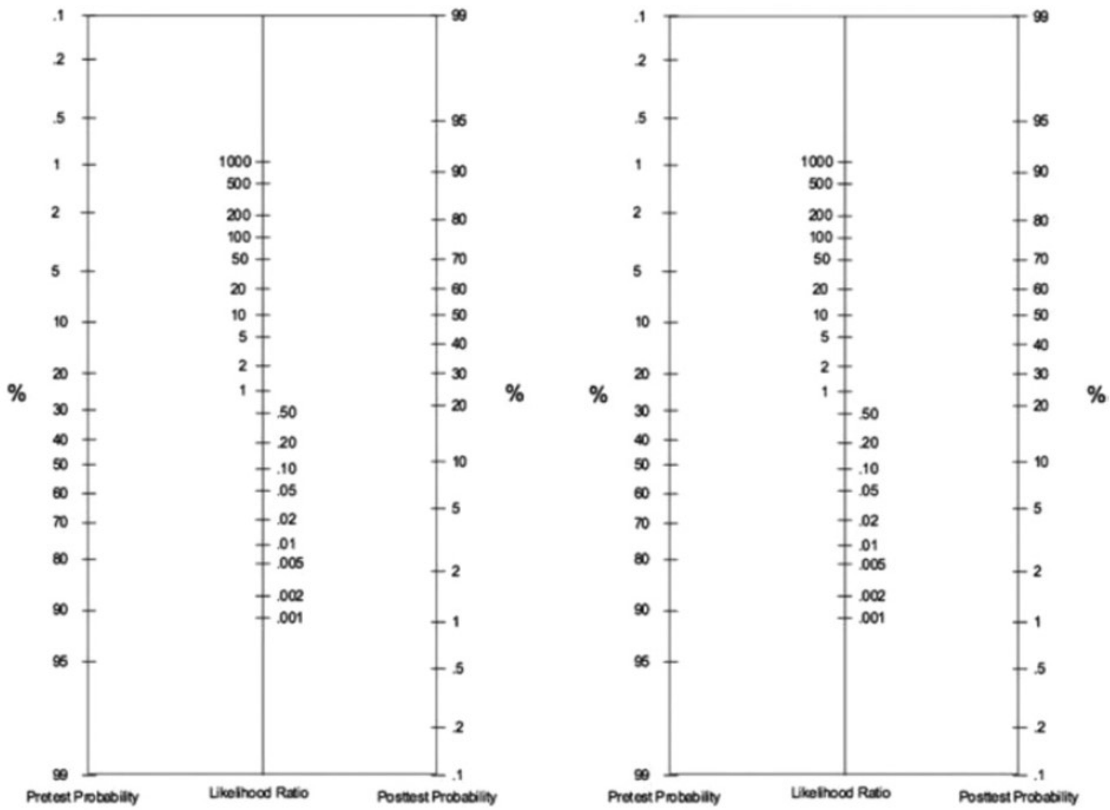
Implementing EBA in Schools

In order to implement EBA, specifically the use of posttest probability, into school practices, it is important to examine the steps of evidence-based assessment and consider how they can be applied. Youngstrom et al. (2015) laid out a series of steps for psychologists in clinical settings to follow: the first is identifying the most common diagnoses in the setting. This step will aid in selecting the appropriate assessments. The second step is to benchmark base rates to anchor evaluations and better inform what issues to prioritize. Base rates can be tallied from data drawn from prior cases and compared to benchmarks from similar establishments and published rates. Third, is to evaluate risk and moderators because risk factors raise "index of suspicion," and the combination of multiple risk factors can elevate probability into "assessment" or possibly "treatment" zones. The fourth step is to synthesize intake instruments into revised posttest probabilities to upgrade the value for formulation and decision-making. These steps are the starting point for implementing evidenced based assessment practices.

Pendergast et al. (2018) provided a translation of these steps to fit a school setting and demonstrated a model of how these steps can be put into practice in schools to address issues with universal screening practices like: inaccuracy of measures, overreliance on single measures, failure to systematically combine information when multiple data points are considered, and failure to incorporate base rates to improve the calibration of decision-making. Their steps begin with identifying the target outcome that is meaningful to school stakeholders by reviewing measurable district goals—in this study it was college readiness based on reading ability as measured by a statewide assessment. Second, they calculated the base rate by identifying the percentage of students in the sample who read below target at the end of the year. However, in practice, schools would usually calculate base rates using extant data from previous years. Third, they identified and evaluated risk factors that were associated with the outcome of interest (curriculum-based measurement scores in reading and behavioral risk). Fourth, they identified likelihood ratios for the predictors and outcomes of interest by using data from a training sample to calculate the correlation between scores on the risk factor measures and scores on the outcome of interest. The authors mentioned again that in practice, districts could calculate the likelihood ratio using data from the prior years. Using the calculated likelihood ratios, posttest probability was calculated using a nomogram, which is a graphing tool used to quickly estimate posterior probability (see Figure 2). Using this approach, 78% of students were correctly identified, which outperforms using the separate screener data alone. Pendergast et al. (2018) provided a concrete example of how posttest probability can be used to effectively and efficiently identify students who may be at risk.

Figure 2

Nomogram



In addition to Pendergast et al. (2018), other studies on screening in schools have included the use of posttest probability (Nelson et al., 2017; Klingbeil et al., 2017, 2019; VanDerHeyden et al., 2017, 2018). These studies examined the diagnostic accuracy of many commonly used screening methods (computer adaptive screeners, CBMs, other published assessments) by calculating posttest probabilities. Results of these studies supported the combination of previous year test scores and computer adaptive or CBM screeners to improve diagnostic accuracy (Klingbeil et al., 2019; Nelson et al., 2017; VanDerHeyden et al., 2017) as well as the use of local or research-derived cut scores (Nelson et al., 2017; Klingbeil et al., 2019). Although these studies provide rudimentary evidence on how schools might implement

EBA (Youngstrom, 2015), by considering base rates and some relevant risk factors (e.g. previous year test performance and screener results), in the calculation of posttest probabilities, evaluation of additional risk factors can be improved.

The Problem

VanDerHeyden (2013) presented the issue that when schools have high screener fail rates, schools add measures for which scores tend to be highly correlated and thus do not meaningfully add to decision accuracy. This issue was echoed in the results of studies by Klingbeil and colleagues (2019) where they found no added benefit of adding more costly screeners (MCOMP and MCAP) to the screening procedure already in place; and by VanDerHeyden and colleagues (2017, 2018) where they found that when schools use multiple measures on one screening occasion (often referred to as "triangulating data"), it was associated with more decision error. These findings beg the question of what then schools are to do when posttest probabilities fall into the test threshold. This is where evaluation of additional risk factors may be useful. Only one of the aforementioned studies included predictor variables other than test or screener scores as likelihood ratios in the calculation of posttest probability. VanDerHeyden et al. (2017) incorporated demographic risk factors such as SES, ELL status, and special education status; although the demographic risk factors were ultimately removed as they did not differentially predict year-end math proficiency in the study once preceding year test scores were entered as a predictor, additional exploration of the impact of diverse, relevant risk factors on posttest probability is warranted. Many of these factors or predictors of achievement are readily available to schools and can provide additional information to the posttest probability without the need to administer another screening test which may conserve time for instruction.

Therefore, to address the problem presented by VanDerHeyden (2013), this review will examine potential factors that may result in more accurate and actionable posttest probabilities and provide an example of how they can be applied in practice.

Socioeconomic Status

The impact that socioeconomic status, and more specifically poverty, has on child development and educational achievement has been well-researched and documented. Studies have found relationships linking poverty in childhood to differential brain development (Hair et al., 2015), lower school readiness, and socioemotional and physical health problems that can undermine educational achievement (Engle & Black, 2008; McLoyd, 1998). These relationships have lasting impacts throughout an individual's school career and into adulthood (Duncan et al., 2012). Kainz (2019) found that students who came from backgrounds of high poverty had lower reading and math scores at the outset of Kindergarten and made lower gains in first grade compared with their non-poor peers. This finding suggests that the readiness and achievement gaps persist as students progress through their education. Further evidence of this is found in studies by Caldas and Bankston (1997) and Gordon and Cui (2018) who have researched the relationship between socioeconomic status (SES) and academic achievement in high school students. The clear link between socioeconomic status and academic outcomes make it an important factor to consider when calculating a posttest probability.

Free and Reduced Lunch (FRL) Status

In educational research, low-income students are typically identified by whether they qualify for the federal free and reduced lunch program (Olszewski-Kubilius & Corwith, 2018). Students are considered eligible for the program if they come from families whose incomes are at or below 130-185% the federal poverty level. Students are also eligible based on their status as a homeless, migrant, runaway, or foster child, or if they are enrolled in federal programs like the Supplemental Nutrition Assistance Program and Head Start (USDA's Food and Nutrition

Service, 2017). It is not an exact measure of poverty as nearly half of students nationwide are eligible for subsidized meals but only a quarter of U.S. children live in poverty; however, it is often the only measure regarding family income readily available in school administrative datasets and has become a proxy measure of socioeconomic status (Michelmore & Dynarski, 2017).

Numerous studies have been conducted that demonstrate the relationship between FRL status and academic achievement both at the individual level and at the district level. Caldas and Bankston (1997), who used participation in FRL to indicate an individual's poverty status, found that poverty status had a direct, independent, statistically significant negative effect on academic achievement for high school students after controlling for other individual-level variables like race and organized activity level. Jones and colleagues (2018) found similar effects of FRL status on standardized test performance for middle school students in a study examining the effects of parent marital status and FRL status on performance on a state test used to gauge college readiness. The results showed statistically significant differences between groups regarding poverty and no main effect or interaction with parent marital status (Jones et al., 2018). Morrissey et al. (2014) also concluded FRL status independently predicts lower grades for elementary school students and is associated with lower standardized test scores.

Additional considerations regarding FRL status

There is a clear increased risk of lower academic achievement for students who are eligible for FRL. Therefore, schools can consider using FRL eligibility as a dichotomized likelihood ratio: students eligible for FRL are at risk, students not eligible are not at risk. However, for schools where a large majority of students qualify for FRL, FRL eligibility alone may not be useful in identifying students for additional support. Because so many students share this risk factor, it would likely provide little predictive information regarding students' future risk of failure. Schools with high percentages of students eligible for FRL may consider alternative criterion to differentiate the level of risk.

First, there is some evidence that shows differential impacts on achievement based on whether the child qualifies for free lunch versus reduced lunch. Morrissey et al. (2014) found that compared to those paying full price for lunch, those receiving free lunch obtained grades that were 18.3% lower whereas those receiving reduced lunch obtained grades that were 6.2% lower. This may be explained by the fact that those who are eligible for free lunch are from more economically disadvantaged homes than those who receive reduced-price lunch. Thus, schools may consider determining risk level based on qualification for free versus reduced-price lunch.

Another potential point for consideration is how long or often a student has been eligible for FRL. In a study on 8^{th} grade math achievement gaps, Michelmore and Dynarski (2017) classified students into categories of economic disadvantage: persistently disadvantaged were students who have been eligible for FRL in every year since kindergarten; transitorily disadvantaged were students who spent some years eligible for FRL; and never disadvantaged were students who never qualified for FRL. Those who were persistently disadvantaged scored an average of .23 standard deviations below those who were transitorily disadvantaged and .69 standard deviations below those who were never disadvantaged. The gap between never disadvantaged and persistently disadvantaged was 41% larger than the gap between never disadvantaged and those disadvantaged at the time of testing. These results indicate the differential impact on academic achievement based on how many years a student has qualified for FRL and may serve as another alternative use of FRL status to determine risk.

Considering the impact economic disadvantage can have on student achievement and the empirical evidence of the relationship between FRL status and academic outcomes, FRL status of a student is a good candidate to consider as a factor included in the calculation of posttest probability. There are also potential options to adjust how to determine cut points for this variable, particularly for schools in which a vast majority of students qualify for FRL. Additionally, FRL status is data most schools have on hand and would be readily available to use with low cost or effort as part of the screening process.

Attendance

When students are absent from school, they miss out on the educational opportunities provided by teacher-led instruction, peer interaction, and other learning activities that prepare them for academic success. The impact that student absenteeism has on achievement has been evinced by numerous studies which cite detrimental outcomes such as poorer test score performance (Aucejo & Romano, 2016), lower math and reading achievement (Gottfried, 2009), and higher likelihood of drop-out (Schoeneberger, 2012). In addition, an increase in absences may exacerbate other academic and sociological risk factors in later years (Gottfried, 2009). Considering the impact that attendance has on student achievement, as well as the accessibility to this data in schools, it may be another useful factor to incorporate when determining posttest probability.

Students might miss a few days for a variety of reasons: personal or family illness, appointments, vacation, or sometimes lack of motivation. On average, students miss around 5 to 6 school days per year which equates to about 3% of the school year (Aucejo & Romano, 2016; Romero & Lee, 2007). In order to utilize attendance data in a posttest probability, it is crucial to identify at what point absences make a significant difference in academic performance and for whom.

Chronic absenteeism

Chronic absenteeism is a term that commonly appears in literature surrounding attendance. Though there is not necessarily a uniform definition across states, researchers typically define chronic absenteeism as missing 10 percent or more of the school year or missing 18 or more school days within a school year (Chang & Davis, 2015). Chronic absenteeism impacts students across all grade levels, and students who are chronically absent in one year are often chronically absent in multiple years (Balfanz & Byrnes, 2012). A national report on chronic absenteeism in early grades (Romero & Lee, 2007) found that students who were chronically absent scored five points lower on academic assessments than students who had average attendance records; these effects often carry over into subsequent school years as demonstrated by a study showing the relationship between chronic absence in Kindergarten and lower academic performance in first grade (Balfanz & Byrnes, 2012). Because there have been numerous national reports and studies illustrating the clear connection chronic absenteeism has with lower academic achievement, chronic absence may be a reliable indicator to use to determine the likelihood ratio for attendance when it is accounted for in the posttest probability.

Additional considerations regarding attendance

The metric of chronic absence (10% of days missed) is a useful starting point, but there are other considerations schools may take into account when determining a cut point. An absence rate of lower than 10 percent can still impact academic outcomes. In addition to labeling chronic absenteeism in their report, Romero and Lee (2007) classified at-risk absentees, who are students that miss 12-18 days in the school year and showed that these early elementary students score an average of 3 points lower on achievement assessments. Another study comparing average number of days absent for high school students who were failing proficiencies in English and math across multiple cohorts found that the average number of absences for those failing either or both proficiencies ranged from 6 to 18 and was almost double those of passing students whose absences ranged between 3 and 8 (Nichols, 2003).

There are many possible circumstances for which schools may choose a lower cut point for classifying at-risk attendance than chronic absence. For one, several studies have noted that the impact of absences on achievement may be exacerbated for students of low socioeconomic backgrounds (Gottfried, 2011; Romero & Lee, 2007). The national dataset from the report by Romero and Lee (2007) showed that the impact of absence on achievement was almost twice as great for poor students. Another reason schools may choose a lower cut point could be dependent on the achievement area of focus. There is evidence that suggests that math achievement is particularly sensitive to number of absences (Balfanz & Byrnes, 2012). Gottfried (2011) found that the negative relationship between missing school and standardized testing achievement is slightly greater for math than for reading; Aucejo (2016) echoes that finding by citing that decreases in absences could increase math scores almost double what the impact on reading scores would be. A third consideration for a revised cut point is absence type. In a study comparing absence impact on achievement after differentiating excused and unexcused absences, Gottfried (2009) found that having a higher proportion of unexcused absences are more likely to experience the negative effects of missing school. Thus, for students with higher ratios of excused absences, the cut point may be higher.

In summation, the clear association between attendance and academic achievement make it a valuable source of additional information to incorporate when calculating a posttest probability. Schools may use chronic absence as the risk determinant or may choose to adjust the cut point depending on contextual considerations. Based on the review of the literature, the cut points should range from 6-18+ days absent and schools or districts can investigate these cut points based on their student data. Because schools regularly collect attendance data, it is an accessible and uncostly addition to the universal screening process.

Behavior

There is an undeniable link between behavior difficulties and poor academic achievement; the co-occurrence of behavior and academic problems is more prevalent than can be explained by chance alone (Hinshaw, 1992). Students with externalizing behavior problems have been found to have lower grades point averages, and higher rates of course failure, retention, and drop out (Wagner, 1995). Researchers have sought to better understand the association between behavior and achievement and the impact of behavior on distal outcomes.

Reinke and colleagues (2008) identified classes of first-grade students with different patterns of academic problems. Using latent class analysis, they distinguished four classes for boys (academic and behavior problems; academic problems only; behavior problems only; and no problem) and three for girls (academic and behavior problems; academic problems only; and no problem). Darney et al. (2013) extended upon this research by investigating the distal outcomes of the same students in twelfth grade. They found that the class with co-occurring academic and behavior problems in the first grade had the greatest risk for negative outcomes including higher likelihood of special education placement, mental health service use, poor academic achievement, and dropout (Darney et al., 2013).

In a separate study, King and colleagues (2015) found similar subgroups of third-grade students but only identified three classes (high degree of academic and some behavioral risk; some academic risk and little behavior risk; and minimal academic and behavior risk). They further validated the classes by comparing these groups' performance on the year end state assessment tests and finding each groups' scores were significantly different from one another: those with a high degree of academic and some behavioral risk performing the lowest and those with minimal academic and behavioral risk performing the highest. Their results support those of Darney et al. (2013) in showing the increased risk that comorbidity of academic and behavior problems can have on achievement. Consequently, behavior is another factor that may contribute to a more accurate posttest probability when predicting academic risk.

Office Discipline Referrals

Office discipline referrals (ODRs) are standardized records of events of problem behavior that occur in schools (Sugai et al., 2000) and can be used as indicators of student behavior problems (McIntosh et al., 2009). ODRs capture a variety of different problem behaviors like physical aggression/fighting, gang affiliation display, defiance, disruption, and more (McIntosh et al., 2010). Schools often use online database systems to collect and report ODR data (McIntosh et al., 2010) making the data easily accessible, which can save schools valuable time and resources, especially compared to administering additional assessments.

Many studies have explored the use of ODR data to predict academic and behavioral outcomes. In a study reporting on an archival review of a sample of 526 students, Tobin and Sugai (1999) found differences in high school outcomes based on how many referrals were received in sixth grade. Results showed that boys with 1-2 ODRs for fighting in sixth grade were on track to graduate in high school but boys with 3 or more ODRs were not. Girls with one ODR for a violent/harassing type incident in sixth grade predicted not being on track to graduate. Similarly, McIntosh et al. (2008) found that students with more referrals had lower average grade point averages. The average grades of students with two or more referrals dropped from fall to spring, whereas average grades of students with up to one referral was stable (McIntosh et al., 2008). Rusby and colleagues (2007) also investigated ODRs as predictors and found that receiving greater levels of referrals in first grade predicted parent and teacher reports of problem behavior at the end of the school year. Overall, the results of these studies illustrate the association between number of ODRs received and academic achievement. It is evident that students who receive more ODRs also tend to have a higher risk for negative outcomes. As such, ODR data can be useful in decision making processes to identify students with increased need for support.

ODR Cut Points

ODR data have been used by schools to determine the level of support required by students in tiered systems of support (Irvin et al., 2004; 2006; McIntosh et al., 2009). To aid this work, researchers have established common cut points that have been subsequently adopted by school personnel; these cut points group students into tiers of support based on number of ODRs received per year: zero to one, two to five, and six or more (McIntosh et al., 2009). These cut points may be useful in determining likelihood ratios as well because there is initial evidence supporting their validity: Walker et al. (2005) found significantly different scores on a social behavior scale between those with zero to one ODRs and those with 2 or more ODRs. The authors chose not to include the third category of 6 or more ODRs due to the small sample of students in that category and lack of significant differences in their social behavior scale scores compared with those in the 2 to 5 ODR group. It is worth noting that the scale used in this study had norms that were at least ten years old at the time. In a more recent study, McIntosh and colleagues (2009) found significant differences in clinical levels of behavioral symptoms and numbers of suspensions between all three groups. The sample size of this study was also small, but findings showed students with 6 or more ODRs had significantly higher levels of externalizing behavior symptoms and suspensions than those with 2 to 5 ODRs who in turn had significantly higher levels of externalizing symptoms and suspensions than those with 0 to 1 ODRs. Based on the results of these studies, having two or more ODRs in one school year appears to increase the risk of negative behavioral outcomes. Given the link between behavioral outcomes and academic achievement, this provides initial evidence to support the use of 2 or more ODRs as a cut point used to estimate the likelihood ratios.

The aforementioned cut points can be used when there is a year's worth of data available, but that may not be the case for every student. To address this issue, researchers have studied ODR growth trajectories to predict which ODR cut point group students will end up in based on ODR type and number of ODRs received in the first few months of the school year. McIntosh et al. (2010) found that receiving two or more ODRs by the end of September was a moderately accurate predictor of receiving 6 ODRs total and became more accurate after each month. By the end of October prediction was highly accurate (McIntosh et al., 2010). They also found that the addition of ODR type enhanced accuracy of prediction with physical aggression and harassment being powerful predictors in middle school and physical aggression and disrespect being moderate predictors in elementary school (McIntosh et al., 2010). Predy et al. (2014) had similar findings in that the most accurate screening results included the type of referral received. Their results also showed that students who received an ODR in any of the first three months of the school year were significantly more likely to have 2-5 ODRs or 6 ODRs by the end of the school year (Predy et al., 2014). They also found that having an ODR for defiance in September was the strongest predictor of having a high total of ODRs at the end of the year. The results of these studies indicate that when year-end ODR totals are not available, receiving one or more ODRs in the first three months of the school year or an ODR for defiance in September can predict risk of receiving 2 or more ODRs by the end of the school year.

There is substantial evidence of the relationship between behavior and academics as well as the association between number of ODRs received and negative academic and behavior outcomes. Therefore, behavior problems as measured by ODRs may be an informative factor to include in the calculation of posttest probability. Receiving two or more ODRs in a school year increases risk of negative outcomes for students, so this could be used as a cut point when evaluating ODR data from the prior school year. When full-year data is not available, ODRs received in the first few months of school can predict year-long outcomes and may be useful in winter screening. However, these cut points should be empirically evaluated. The potential predictive value associated with ODR data and the fact that ODRs are widely collected in schools make it another relevant factor to explore for use in determining posttest probability.

Application Example and Future Directions

To illustrate how these recommended factors could be used to calculate a posttest probability in practice, an example is provided using hypothetical data. This example will walk through each step of calculating a revised posttest probability including the factors discussed for a hypothetical student, Harry.

Harry attends a school where the rate of students scoring below target on the state reading achievement test is 35% [pretest probability]. Harry passed the previous year's state reading test but failed the fall screener which resulted in a posttest probability of 37%. Following VanDerHeyden's (2013) suggested thresholds, additional assessment would need to be conducted to determine whether Harry should receive intervention. This is where likelihood ratios derived from the factors described above could be useful. When updating the posttest probability based on new information, the first step is to use the posttest probability from the last step as the revised pretest probability in the calculation. In this example, Harry's revised pre-test probability was .37 after applying the results from the state reading achievement test from the last year and the fall screener.

Free and Reduced Lunch Status. Thirty percent of students at Harry's school qualify for Free or Reduced Lunch and the decision-making team determined that using FRL status to calculate posttest probability was appropriate as sensitivity was high. The corresponding calculations are shown in Table 1. Harry qualifies for FRL and therefore is considered at-risk. Because he is at-risk, a positive likelihood ratio is calculated using the sensitivity and specificity, which are the percentage of students that were not proficient on the state test who were eligible for FRL and the percentage of students that were proficient on the state test who were not eligible for FRL respectively. Then the pretest odds were calculated using the posttest probability as the revised pretest probability. The pretest odds were multiplied with the positive likelihood ratio to get the posttest odds which were then used to calculate the revised posttest probability. After factoring in FRL status, Harry's posttest probability is 47% meaning additional assessment still needs to be done.

Table 1.

Calculation of Revised Posttest Probability with FRL status

<table>
<tr><th rowspan="2">FRL eligibility</th><th colspan="6">State Test</th></tr>
<tr><th>Proficient (n (%))</th><th>Not proficient (n (%))</th><th>Likelihood Ratio</th><th>Pretest Odds</th><th>Posttest Odds</th><th>Revised Posttest Probability</th></tr>
<tr><td>Not Eligible</td><td>84 (40)</td><td>9 (10)</td><td>(1-.9)/.4=.25</td><td rowspan="2">.37/(1-.37) =.59</td><td rowspan="2">.59×1.5 =.89</td><td rowspan="2">.89/(1+.89) =.47</td></tr>
<tr><td>Eligible</td><td>126 (60)</td><td>81 (90)</td><td>.9/(1- 4)=1.5</td></tr>
</table>

Note. Pretest odds are calculated using the posttest probability as the revised pretest probability.

Attendance. In this hypothetical example, Harry's school has determined that their cut point for at-risk attendance was 9 absences or 5% of days missed. They determined this number because it had 90% sensitivity and 70% specificity. In the previous school year Harry had missed 6% of days and was determined at-risk. Now the revised posttest probability would be used to calculate the new pretest odds. Calculations are shown in Table 2 and follow the same procedure as described with FRL status. The revised posttest probability is now 73% which is above the threshold VanDerHeyden (2013) suggested for intervention.

Table 2.

Calculation of Revised Posttest Probabilities with Days Absent

<table>
<tr><th rowspan="2">Days Absent > 5%</th><th colspan="6">State Test</th></tr>
<tr><th>Proficient (n (%))</th><th>Not proficient (n (%))</th><th>Likelihood Ratio</th><th>Pretest Odds</th><th>Posttest Odds</th><th>Revised Posttest Probability</th></tr>
<tr><td>No</td><td>122 (70)</td><td>17 (10)</td><td>(1-.9)/.7=.14</td><td rowspan="2">.47/(1-.47) = .89</td><td rowspan="2">.89×3 =2.67</td><td rowspan="2">2.67/(1+2.67) =.73</td></tr>
<tr><td>Yes</td><td>52 (30)</td><td>113 (90)</td><td>.9/(1-.7) =3</td></tr>
</table>

Note. Pretest odds are calculated using the posttest probability as the revised pretest probability.

This example illustrates how a school can apply the use of posttest probability to the decision-making process within the context of multi-tiered systems of support. To date, the utility of these variables in the estimation of posttest probability is conjecture and should be validated by research.

Conclusion and Future Directions

According to prominent researchers in this area (e.g., VanDerHeyden & Burns, 2018), the use of posttest probability can help schools move toward a more systematic, actuarial approach to assessment that is evidence based. This may help address some of the issues with current screening practices in schools like over- or under-identification of students at-risk through incorporating local base rates of failure in the assessment of risk. Posttest probabilities are also actionable metrics that, when used in a threshold model, can inform decision-making in schools.

When a student's posttest probability falls into the threshold that indicates the need for assessment, schools could choose to collect further assessment information using other screening measures or diagnostic tests. However, in practice, schools that use multiple screening measures tend to use measures that are highly correlated and do not meaningfully add to the accuracy of decision (VanDerHeyden, 2013). Thus, this literature review has sought to examine alternative predictors of academic risk that could be useful for updating posttest probabilities without the need to administer additional assessments that can ultimately lead to decision error. Socioeconomic status, as measured by FRL eligibility, attendance, and behavior, as measured by office discipline referrals, are all data that schools already collect and can be easily accessed by the school's decision-making team without adding cost or labor to the assessment process.

Empirical evaluation of the use of these three risk factors in posttest probabilities is a rich area for future research. Research on the use of posttest probabilities in schools is limited, and research on the use of posttest probabilities including demographic data is even more so. It should be noted that the factors covered in this review are not the only factors that could be effective in achieving more accurate posttest probabilities, and other student characteristics like English learner and special education status, among others, may also be areas worthy of investigation. Researchers can look into the utility and validity of school or district cut scores derived from data from previous years for each of the factors. In addition, researchers can also investigate the validity of modified cut scores or criteria for different contexts; For example, looking at using persistently disadvantaged and transitorily disadvantaged as the risk determinant for schools with high poverty or looking at whether using ODR data from the first few months of school is valid for winter screening.

In conclusion, there is much to be evaluated in terms of the use of posttest probability in schools as well as which factors should be considered to calculate the most accurate probability. Nevertheless, posttest probability and the addition of demographic variables in its calculation offer a promising way to use universal screening data that is aligned with evidence-based assessment practices. If future research validates the use of such factors in an evidence-based assessment approach to screening, it will enable school psychologists and other members of a school's problem-solving team to make better, more informed decisions about their students while conserving resources for instruction.

References

Balfanz, R., & Byrnes, V. (2012). The importance of being in school: A report on absenteeism in the nation's public schools. *The Education Digest*, *78*(2), 4. https://coseboc.org/sites/coseboc.org/files/assets/ChronicAbsenteeismReport.pdf

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). Evaluation of Response to Intervention Practices for Elementary School Reading. NCEE 2016-4000. *National Center for Education Evaluation and Regional Assistance*. https://files.eric.ed.gov/fulltext/ED560820.pdf

Caldas, S. J., & Bankston, C. (1997). Effect of school population socioeconomic status on individual academic achievement. *The Journal of Educational Research*, *90*(5), 269-277. https://doi.org/10.1080/00220671.1997.10544583

Chang, H. N., & Davis, R. (2015). Mapping the early attendance gap. *San Francisco, CA: Attendance Works*. https://coseboc.org/sites/coseboc.org/files/assets/ChronicAbsenteeismReport.pdf

Clemens, N. H., Keller-Margulis, M. A., Scholten, T., & Yoon, M. (2016). Screening assessment within a multi-tiered system of support: Current practices, advances, and next steps. In *Handbook of response to intervention* (pp. 187-213). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7568-3_12

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of school psychology*, *51*(1), 117-128. https://doi.org/10.1016/j.jsp.2012.09.005

Duncan, G. J., Magnuson, K., Kalil, A., & Ziol-Guest, K. (2012). The importance of early childhood poverty. *Social Indicators Research*, *108*(1), 87-98. https://doi.org/10.1007/s11205-011-9867-9

Engle, P. L., & Black, M. M. (2008). The effect of poverty on child development and educational outcomes. *Annals of the New York Academy of Sciences*, *1136*, 243. https://doi.org/10.1196/annals.1425.023.

Gordon, M. S., & Cui, M. (2018). The intersection of race and community poverty and its effects on adolescents' academic achievement. *Youth & Society*, *50*(7), 947-965. https://doi.org/10.1177/0044118X16646590

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*(2), 117-135. https://doi.org/10.1016/j.jsp.2006.05.005

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of child poverty, brain development, and academic achievement. *JAMA pediatrics*, *169*(9), 822-829. https://doi.org/10.1001/jamapediatrics.2015.1475

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. *Psychological bulletin*, *111*(1), 127. https://doi.org/10.1037/0033-2909.111.1.127

Irvin, L. K., Horner, R. H., Ingram, K., Todd, A. W., Sugai, G., Sampson, N. K., & Boland, J. B. (2006). Using office discipline referral data for decision making about student behavior in elementary and middle schools: An empirical evaluation of validity. *Journal of*

Positive Behavior Interventions, *8*(1), 10-23. https://doi.org/10.1177/10983007060080010301

Irvin, L. K., Tobin, T. J., Sprague, J. R., Sugai, G., & Vincent, C. G. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status and effects of school-wide behavioral interventions. *Journal of Positive Behavior Interventions*, *6*(3), 131-147. https://doi.org/10.1177/10983007040060030201

Jenkins, J. R., Schiller, E., Blackorby, J., Thayer, S. K., & Tilly, W. D. (2013). Responsiveness to intervention in reading: Architecture and practices. *Learning Disability Quarterly*, *36*(1), 36-46. https://doi.org/10.1177/0731948712464963

Jones, K., Wilson, R., Clark, L., & Dunham, M. (2018). Poverty and parent marital status influences on student achievement. *Educational Research Quarterly*, *42*(1), 62-80.

Kainz, K. (2019). Early academic gaps and Title I programming in high poverty, high minority schools. *Early Childhood Research Quarterly*, *47*, 159-168. https://doi.org/10.1016/j.ecresq.2018.08.012

Kilgus, S. P., & Eklund, K. R. (2016, March). Consideration of base rates within universal screening for behavioral and emotional risk: A novel procedural framework. In *School Psychology Forum* (Vol. 10, No. 1). https://www.researchgate.net/profile/Stephen_Kilgus/publication/299437499_Consideration_of_Base_Rates_Within_Universal_Screening_for_Behavioral_and_Emotional_Risk_A_Novel_Procedural_Framework/links/56f6ed3708ae81582bf2fb71.pdf

King, K. R., Lembke, E. S., & Reinke, W. M. (2016). Using latent class analysis to identify academic and behavioral risk status in elementary students. *School Psychology Quarterly*, *31*(1), 43. https://doi.org/10.1037/spq0000111

Klingbeil, D. A., Maurice, S. A., Van Norman, E. R., Nelson, P. M., Birr, C., Hanrahan, A. R., ... & Lopez, A. L. (2019). Improving Mathematics Screening in Middle School. *School Psychology Review*, *48*(4), 383-398. DOI: 10.17105/SPR-2018-0084.V48-4

Klingbeil, D. A., Nelson, P. M., Van Norman, E. R., & Birr, C. (2017). Diagnostic accuracy of multivariate universal screening procedures for reading in upper elementary grades. *Remedial and Special Education*, *38*(5), 308-320. https://doi.org/10.1177/0741932517697446

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American psychologist*, *53*(2), 185. https://doi.org/10.1037/0003-066X.53.2.185

McIntosh, K., Brigid Flannery, K., Sugai, G., Braun, D. H., & Cochrane, K. L. (2008). Relationships between academics and problem behavior in the transition from middle school to high school. *Journal of Positive Behavior Interventions*, *10*(4), 243-255. https://doi.org/10.1177/1098300708318961

McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009). Concurrent validity of office discipline referrals and cut points used in schoolwide positive behavior support. *Behavioral Disorders*, *34*(2), 100-113. https://doi.org/10.1177/019874290903400204

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing research-based trajectories of office discipline referrals for individual students. *School Psychology Review*, *39*(3), 380-394.

Mellard, D. F., McKnight, M., & Woods, K. (2009). Response to intervention screening and progress-monitoring practices in 41 local schools. *Learning Disabilities Research & Practice*, *24*(4), 186-195. https://doi.org/10.1111/j.1540-5826.2009.00292.x

Michelmore, K., & Dynarski, S. (2017). The gap within the gap: Using longitudinal data to understand income differences in educational outcomes. *AERA Open*, *3*(1), 2332858417692958. https://doi.org/10.1177/2332858417692958

Morrissey, T. W., Hutchison, L., & Winsler, A. (2014). Family income, school attendance, and academic achievement in elementary school. *Developmental psychology*, *50*(3), 741. https://doi.org/10.1037/a0033848

Nelson, P. M., Van Norman, E. R., & VanDerHeyden, A. (2017). Reduce, reuse, recycle: The longitudinal value of local cut scores using state test data. *Journal of Psychoeducational Assessment*, *35*(7), 683-694. https://doi.org/10.1177/0734282916658567

Pauker, S. G., & Kassirer, J. P. (1980). The threshold approach to clinical decision making. *New England Journal of Medicine*, *302*(20), 1109-1117. https://doi.org/10.1056/NEJM198005153022003

Predy, L., McIntosh, K., & Frank, J. L. (2014). Utility of number and type of office discipline referrals in predicting chronic problem behavior in middle schools. *School Psychology Review*, *43*(4), 472-489. DOI: 10.17105/SPR-13-0043.1

Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: Co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology*, *36*(5), 759-770. https://doi.org/10.1007/s10802-007-9208-2

Romero, M., & Lee, Y. S. (2007). A national portrait of chronic absenteeism in the early grades. https://doi.org/10.7916/D89C7650

Rusby, J. C., Taylor, T. K., & Foster, E. M. (2007). A descriptive study of school discipline referrals in first grade. *Psychology in the Schools*, *44*(4), 333-350. https://doi.org/10.1002/pits.20226

Straus, S. E., Glasziou, P., Richardson, W. S., & Haynes, R. B. (2018). *Evidence-Based Medicine E-Book: How to Practice and Teach EBM*. Elsevier Health Sciences. https://doi.org/10.1111/j.1540-5826.2009.00292.x

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of emotional and behavioral disorders*, *8*(2), 94-101. https://doi.org/10.1177/106342660000800205

Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, *7*(1), 40-53. https://doi.org/10.1177/106342669900700105

USDA's Food and Nutrition Service. (2017). *The National School Lunch Program Fact Sheet.* United States Department of Agriculture. https://fns-prod.azureedge.net/sites/default/files/resource-files/NSLPFactSheet.pdf

VanDerHeyden, A. M. (2013). Universal screening may not be for everyone: Using a threshold model as a smarter way to determine risk. *School Psychology Review*, *42*(4), 402-414. DOI: 10.1080/02796015.2013.12087462

VanDerHeyden, A. M., & Burns, M. K. (2018). Improving decision making in school psychology: Making a difference in the lives of students, not just a prediction about their lives. *School Psychology Review*, *47*(4), 385-395. DOI: 10.17105/SPR-2018-0042.V47-4

VanDerHeyden, A. M., Burns, M. K., & Bonifay, W. (2018). Is more screening better? The relationship between frequent screening, accurate decisions, and reading proficiency. *School Psychology Review*, *47*(1), 62-82. DOI: 10.17105/SPR-2017-0017.V47-1

VanDerHeyden, A. M., Codding, R. S., & Martin, R. (2017). Relative value of common screening measures in mathematics. *School Psychology Review*, *46*(1), 65-87. DOI: 10.1080/02796015.2017.12087608

Wagner, M. M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *The Future of Children*, 90-112. https://doi.org/10.2307/1602359

Walker, B., Cheney, D., Stage, S., Blum, C., & Horner, R. H. (2005). Schoolwide screening and positive behavior supports: Identifying and supporting students at risk for school failure. *Journal of Positive Behavior Interventions*, *7*(4), 194-204. https://doi.org/10.1177/10983007050070040101

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology*, *42*(1), 139-159. https://doi.org/10.1080/15374416.2012.736358

Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2015). Clinical guide to the evidence-based assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*, *22*(1), 20-35. https://doi.org/10.1016/j.cbpra.2013.12.005